

## **JP9065303**

Publication Title:

**VIDEO SOUND SIGNAL SYNCHRONIZATION METHOD AND ITS DEVICE**

Abstract:

Abstract of JP9065303

**PROBLEM TO BE SOLVED:** To synchronize a sound output with a reproduced video image by configuring the device such that the sound output is synchronously with a reproduction output of a moving image packet. **SOLUTION:** A video transmitter 10 adds a time code to sound data and moving image data respectively and outputs the result to a network 18. A moving image data receiver 22 of a video reproduction terminal equipment 20 receives moving image data and a time code from the network 18 and a sound data receiver 24 receives the sound data and the time code. The received moving image data are fed to a monitor display device 30 via a frame buffer 28 and a video image is displayed on a screen. The receiver 24 stores temporarily the received sound data to sound output buffer 32. A sound output timing adjustment device 26 inserts or deletes a silence period in the sound output buffer 32 so that a corresponding sound signal is outputted from a speaker 34 to a frame for video display according to a time code from receivers 22, 24 and buffer consumed amount information of the sound output buffer 32.

Data supplied from the esp@cenet database - Worldwide

-----  
Courtesy of <http://v3.espacenet.com>

# BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-65303

(43) 公開日 平成9年(1997)3月7日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 N 7/10 5/93			H 0 4 N 7/10 5/93	E G

審査請求 未請求 請求項の数 8 O L (全 6 頁)

(21) 出願番号 特願平7-218837

(22) 出願日 平成7年(1995)8月28日

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(72) 発明者 岡崎 洋

東京都大田区下丸子3丁目30番2号キヤノン株式会社内

(72) 発明者 佐藤 宏明

東京都大田区下丸子3丁目30番2号キヤノン株式会社内

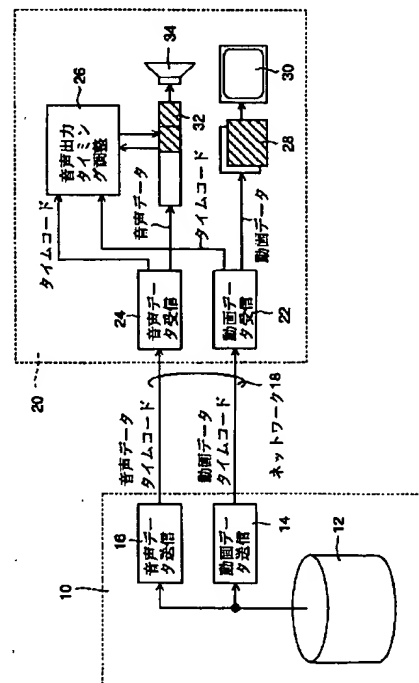
(74) 代理人 弁理士 田中 常雄

(54) 【発明の名称】 映像音声同期方法及び装置

(57) 【要約】

【課題】 音声出力と映像再生を同期させる。

【解決手段】 ビデオ送信装置10は、音声データと動画データの夫々にタイムコードを付加してネットワーク18に出力する。ビデオ再生端末装置20の動画データ受信装置22がネットワーク18からの動画データとタイムコードを受信し、音声データ受信装置24が音声データとタイムコードを受信する。受信された動画データはフレーム・バッファ28を介してモニタ・ディスプレイ30に印加され、その画面上に映像表示される。受信装置24は、受信した音声データを音声出力バッファ32に一時格納する。音声出力タイミング調整装置26は、受信装置22、24からのタイムコード及び音声出力バッファ32のバッファ使用量情報に従い、映像表示するフレームに該当する音声が入力されるように、音声出力バッファ32上で無音声区間を挿入又は削除する。



## 【特許請求の範囲】

【請求項1】 蓄積されているビデオ・データの動画成分及び音声成分を、それぞれの時間関係を特定する時間情報を付加して独立にネットワークに出力する出力ステップと、

当該ネットワークから動画成分及びその時間情報を受信する動画受信ステップと、

当該ネットワークから音声成分及びその時間情報を受信する音声受信ステップと、

当該音声受信ステップで受信された音声成分を音声出力バッファに一時記憶する一時記憶ステップと、

当該動画受信ステップで受信された動画情報を映像表示する映像表示ステップと、

受信された動画成分の時間情報及び音声成分の時間情報を比較し、受信された動画成分の映像出力と同期するように、当該音声出力バッファの記憶データを調整する調整ステップと、

当該音声出力バッファから出力されるデータを音声出力する音声出力ステップとからなることを特徴とする映像音声同期方法。

【請求項2】 上記出力ステップが、上記音声成分を上記動画成分より所定期間先立ってネットワークに出力する請求項1に記載の映像音声同期方法。

【請求項3】 上記ビデオ・データが、上記動画成分のフレーム単位で蓄積されている請求項1又は2に記載の映像音声同期方法。

【請求項4】 上記調節ステップが、音声再生が動画再生より早い場合に、上記音声出力バッファ上で無音声区間を同期ずれ時間分長くし、音声再生が動画再生より遅い場合に、無音声区間を同期ずれ時間分短くする請求項1乃至3の何れか1項に記載の映像音声同期方法。

【請求項5】 蓄積されているビデオ・データの動画成分及び音声成分を、それぞれの時間関係を特定する時間情報を付加して独立にネットワークに出力する出力装置と、当該出力装置から出力される動画成分及び音声成分をそれぞれ映像表示及び音声出力する再生装置とからなる伝送システムにおいて、当該再生装置の映像表示と音声出力を同期させる映像音声同期装置であって、受信された音声成分を一時記憶する音声出力バッファと、受信された動画成分の時間情報と音声成分の時間情報の比較結果に従い、当該音声出力バッファの記憶データを調整する音声出力タイミング調整手段とからなることを特徴とする映像音声同期装置。

【請求項6】 上記出力装置が、上記音声成分を上記動画成分より所定期間先立って上記ネットワークに出力する請求項5に記載の映像音声同期装置。

【請求項7】 上記ビデオ・データが、上記動画成分のフレーム単位で蓄積されている請求項5又は6に記載の映像音声同期装置。

【請求項8】 上記調整手段が、音声再生が動画再生よ

り早い場合に、上記音声出力バッファ上で無音声区間を同期ずれ時間分長くし、音声再生が動画再生より遅い場合に、無音声区間を同期ずれ時間分短くする請求項5乃至7の何れか1項に記載の映像音声同期装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、映像音声同期方法及び装置に関し、より具体的には、遠隔ディジタル・ビデオ再生システムのように遠隔地に映像と音声を送送する伝送システムにおいて、映像と音声の再生を同期させる方法及び装置に関する。

## 【0002】

【従来の技術】従来、ネットワークを介してビデオ・データを伝送し、遠隔地で再生する遠隔ビデオ再生システムでは、同じ時刻における音声データと動画データを同一パケットに多重配置して送信し、受信側では、同一パケットに含まれる音声データと動画データを同時に再生出力する方式が採用されている。

## 【0003】

【発明が解決しようとする課題】一般的なワークステーションやパーソナル・コンピュータにも、音声や動画をリアルタイムに再生処理する処理系が普通に組み込まれるようになった。そのような状況では、独立に開発された音声処理系と動画処理系を同時に動作させるようにして、比較的安価に遠隔ビデオ再生システムを実現できる。

【0004】図2は、その遠隔ビデオ再生システム概略構成ブロック図を示す。ビデオ送信装置110は、ビデオ・データを格納する大容量記憶装置（例えばハードディスク装置又は光ディスク装置など）112を具備し、動画データ送信装置114が記憶装置112から動画データを読み出してネットワーク118に出力し、音声データ送信装置116が記憶装置112から音声データを読み出してネットワーク118に出力する。

【0005】ビデオ再生端末装置120の動画データ受信装置122がネットワーク118からの動画データを受信し、画像データを復元する。復元された画像データはフレーム・バッファ124を介してモニタ・ディスプレイ126に印加され、モニタ・ディスプレイ126の画面上に映像表示される。また、ビデオ再生端末装置120の音声データ受信装置128は、ネットワーク118からの音声データを受信し、音声データを復元する。復元された音声データは音声出力バッファ130を介してスピーカ132に印加され、スピーカ132から音声出力される。

【0006】図2に示すようなシステム構成でも、ネットワークでの一定でない伝送遅延等を考慮すると、何らかの方法で、動画の再生出力と音声の再生出力との間に同期をとる必要がある。

【0007】先に説明したような、音声データと動画デ

ータを同一パケットに多重配置する構成を組み入れようとする、ネットワーク伝送部分を中心にかなり大規模な修整をプログラムに施す必要が生じる。

【0008】本発明は、ネットワーク伝送部分における修整を最小限にして、映像と音声の同期再生を可能にする映像音声同期方法及び装置を提示することを目的とする。

【0009】

【課題を解決するための手段】本発明では、送信側で、音声パケットと動画パケットのそれぞれにタイムコード情報を埋め込む。受信側では、音声パケットのタイムコードと動画パケットのタイムコードを比較し、動画パケットの再生出力に音声出力が同期するように、音声データの無音声区間を長くしたり、短くしたりする。無音声区間の削除又は挿入は、音声出力バッファ上で行なわれる。これにより、プログラムの修正をなるべく最小にとどめて、音声出力と動画出力を同期させることができる。

【0010】

【発明の実施の形態】以下、図面を参照して、本発明の一実施の形態を詳細に説明する。

【0011】図1は、本発明の一実施例の概略構成ブロック図を示す。10は、ビデオ・データをハード・ディスク装置又は光ディスク装置などの大容量記憶装置12に格納するビデオ送信装置であり、格納されるビデオ・データを、ネットワークを介して自在に読み出させるようになっている。ビデオ・データの動画部分と音声部分は、動画のフレーム単位で記憶装置12に格納される。即ち、動画像がフレーム単位で圧縮され、そのフレーム時間分の音声データが付加された形で記録されている。図3は、ビデオ・データの記録フォーマットを示す。

【0012】ビデオ送信装置10の動画データ送信装置14は、記憶装置12に記憶される動画データを読み出し、そのフレームのタイムコード（最初のフレームからの経過時間情報）と共にネットワーク18に出力する。また、ビデオ送信装置10の音声データ送信装置16は、記憶装置12に記憶される音声データを読み出し、動画データ送信装置16からネットワーク18に出力される動画フレームと同じフレームの音声データを、そのフレームのタイムコードと共にネットワーク18に出力する。動画データ及び音声データと一緒に伝送されるタイムコードにより、同じフレームの画像と音声とが識別可能になる。

【0013】ビデオ再生端末装置20の動画データ受信装置22がネットワーク18からの動画データとタイムコードを受信し、音声データ受信装置24は、ネットワーク18からの音声データとタイムコードを受信する。受信装置22、24が受信したタイムコードは音声出力タイミング調整装置26に引き渡される。動画データ受信装置22は、受信した動画データから各フレームの画

像データを復元し、復元された画像データはフレーム・バッファ28を介してモニタ・ディスプレイ30に印加され、モニタ・ディスプレイ30の画面上に映像表示される。なお、フレーム・バッファ28は2フレーム分かなり、一方のフレームが動画データ受信装置22からの画像データの書き込み用、他方のフレームがモニタ・ディスプレイ30への表示用となり、フレーム毎に書き込み用と表示用が切り換えられるようになっている。

【0014】音声データ受信装置24は、受信した音声データからデジタル音声信号を復元し、復元されたデジタル音声信号は音声出力バッファ32に一時格納される。詳細は後述するが、音声出力タイミング調整装置26は、受信装置22、24からのタイムコードを比較すると共に、音声出力バッファ32からのバッファ使用量情報に従い、映像表示するフレームに該当する音声をスピーカ34から出力するように、即ち、音声出力が映像出力と同期するように、音声出力バッファ32を制御する。

【0015】ビデオ送信装置10及びビデオ再生端末装置20は、図4に示すような基本構成のワークステーションにより実現できる。40は、米国Sun Microsystems社のワークステーションSPARC station、42は、Parallax社製のJPEG圧縮伸長機能付きビデオ・ボードPower Videoである。ビデオ送信装置10及びビデオ再生端末装置20として、それぞれに必要なソフトウェアがインストールされる。

【0016】ビデオ送信装置10は、ネットワーク18を介して何れかのビデオ再生端末装置20から再生要求を受信すると、送信装置14、16がそれぞれ、動画データ及び音声データを記憶装置12から読み出し、ネットワーク18への出力を開始する。

【0017】動画データ送信装置14は、記憶装置12から、要求されたビデオ・データの動画データ・フレームを順次読み込み、そのフレーム・レートでネットワーク18に送信する。送信装置18は、動画データの送信に際して、 $i$  番目のフレームに対し  $(i-1)/fp$  (秒) のタイムコードを付加する。 $fp$  は、記憶装置12に蓄積されているデジタル・ビデオ・データのフレーム・レートを表わす。

【0018】音声データ送信装置16は、記憶装置12から音声データ・フレームを $k$  個先読みし、フレームの抜けなく連続性を保ってネットワーク18に送信する。装置16は、動画の場合と同様に、 $i$  番目のフレームに対し  $(i-1)/fp$  (秒) のタイムコードを付加して送信する。なお、 $k$  は3乃至5程度である。

【0019】ビデオ再生端末装置20における映像と音声の同期化動作を説明する。動画データ受信装置22はネットワーク18上を伝送する動画データとこれに付加されたタイムコードを受信し、音声データ受信装置24

はネットワーク18上を伝送する音声データとこれに付加されたタイムコードを受信する。動画データ受信装置22は、受信した動画データから各フレームの画像データを復元し、フレーム・バッファ28に書き込む。先に説明したように、フレーム・バッファ28に書き込まれた画像データは、逐次、読み出され、モニタ・ディスプレイ30に印加されて、モニタ・ディスプレイ30の画面上に映像表示される。音声データ受信装置24は、受信した音声データからデジタル音声信号を復元し、音声出力バッファ32に書き込む。

【0020】受信装置22、24はまた、受信したタイムコードを音声出力タイミング調整装置26に供給する。音声出力タイミング調整装置26には音声出力バッファ32からバッファ使用量の情報が供給されている。音声出力タイミング調整装置26は、まず、バッファ使用量を音声データのサンプリング・レートで除算して、どの程度の時間後に、音声出力バッファ32から読み出されてスピーカ34から音声出力されるかを計算する。即ち、

$T1 = (\text{バッファ使用量}) / (\text{サンプリング・レート})$ とする。サンプリング・レートは $\mu$ -LAW形式の音声データでは、8,000 bytes/secとなる。

【0021】更に、映像出力に音声出力を合わせるために、受信した音声データのタイムコードと受信した動画データのタイムコードの差 $T2$ を算出する。即ち、 $T2 = (\text{音声データのタイムコード}) - (\text{動画データのタイムコード})$

とする。これは、動画データ受信装置22からのタイムコードを基準とするのが利用的だからである。

【0022】 $T1$ と $T2$ の差がある閾値 $R$ (本実施例では、0.1乃至0.4程度である。)以上であれば、音声出力タイミング調整装置26は、音声出力バッファ32中の音声データのデータ長を、なるべく実際の視聴時に音切れ等が発生しないように調整し、これにより、音声出力を映像出力に同期させる。

【0023】 $T1 - T2 > R$ のときには、下記式で得られる $L$ (バイト)のデータを音声出力バッファ32から取り除く。即ち、

$$L = (T1 - T2) \times (\text{サンプリング・レート}) \\ = (T1 - T2) \times 8,000$$

音声出力バッファ32からデータの一部を取り除く方法として、以下の2方式が考えられる。第1の方法では、音声出力バッファ32中の最初又は最後の $L$ (バイト)を単純に取り除く。第2の方法では、音声出力バッファ32中の無音区間を認識し、そこから $L$ (バイト)を取り除く。第2の方法の方が好ましいのは明らかである。

【0024】第2の方法をより詳細に説明する。音声データをまず、 $s$ サンプリング単位に分割し、1単位区間の $s$ 個の各サンプリング値の分散をとる。この際、サンプリング値は、 $\mu$ -LAW等の符号化された音声デ

ータを扱う場合には、復号化した値を利用する。分散が小さい単位区間は無音声区間候補となり、その単位区間が $n$ 個連続することで、その区間を無音声区間と判断する。音声出力バッファ32中に複数の無音声区間があると、その区間の長さの比によって、取り除くバイト数を配分し、各無音声区間を間引いて短くする。但し、取り除けるバイト数の上限を無音声区間の総計の半分までとしておく。本実施例では、例えば、 $s$ を200乃至400程度、 $n$ を3乃至5程度とした。

【0025】 $T2 - T1 \geq R$ のとき、下記式で得られる $L$ (バイト)の無音声データを音声出力バッファ32中に埋め込む。即ち、

$$L = (T2 - T1) \times (\text{サンプリング・レート}) \\ = (T2 - T1) \times 8,000$$

音声出力バッファ32中に無音声区間を挿入する方法として、以下の2つの方法が考えられる。第1の方法では、音声出力バッファ32の最初又は最後に $L$ (バイト)の無音データを挿入する。第2の方法では、音声出力バッファ32中の無音声区間を認識し、そこに $L$ (バイト)の無音データを挿入する。第2の方法の方が優れているのは明らかである。第2の方法では、無音声データを除去する場合の第2の方法で説明したのと同様の方法で無音声区間を検出する。音声出力バッファ32中に複数の無音声区間があると、その区間の長さの比によって、挿入するバイト数を配分し、各無音声区間を補間して長くする。但し、挿入可能なバイト数の上限を無音声区間の総計の半分までとしておく。

【0026】このように、音声パケット及び動画パケットのそれぞれに同期のための時間情報を埋め込み、受信側の音声出力バッファ32中の無音声区間を上述のように調整することで、音声出力を映像出力に同期させることができる。この結果、音声系と動画系が独立して動作している遠隔デジタル・ビデオ再生システムで、プログラムの修正をなるべく最小にとどめて、音声と動画を同期出力させることができる。

【0027】

【発明の効果】以上の説明から容易に理解できるように、本発明によれば、音声と動画を論理的に別々に伝送する伝送システムにおいて、音声出力と画像出力を互いに同期させることができる。動画と音声を同一のパケットで伝送する訳ではないので、映像・音声の伝送のためのプログラム部分の修正を不要又は最小限で済む。

【図面の簡単な説明】

【図1】 本発明の一実施例の概略構成ブロック図である。

【図2】 従来例の概略構成ブロック図である。

【図3】 記憶装置12に記憶されるビデオ・データのフォーマットである。

【図4】 ビデオ送信装置10及びビデオ再生端末装置20を実現するワークステーションの概略構成ブロック

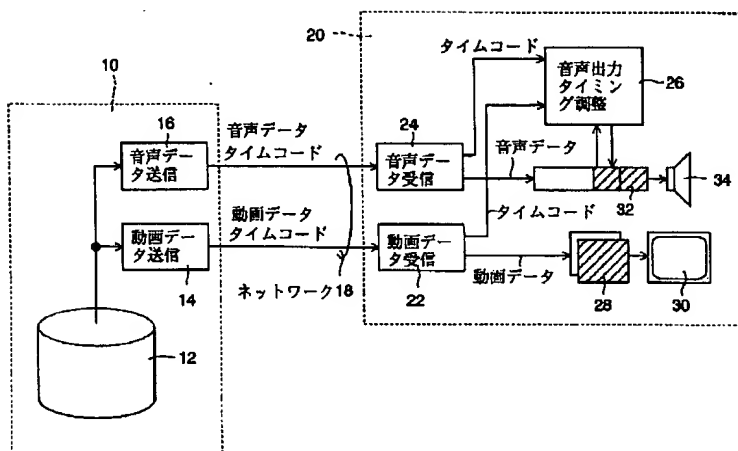
図である。

【符号の説明】

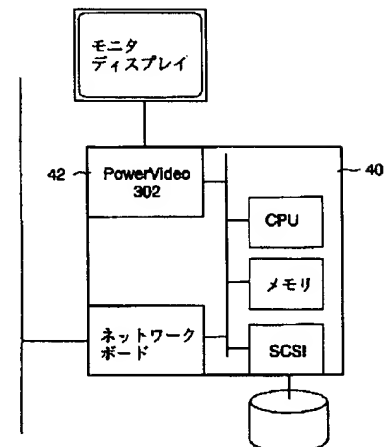
- 10：ビデオ送信装置
- 12：大容量記憶装置
- 14：動画データ送信装置
- 16：音声データ送信装置
- 18：ネットワーク
- 20：ビデオ再生端末装置
- 22：動画データ受信装置
- 24：音声データ受信装置
- 26：音声出力タイミング調整装置
- 28：フレーム・バッファ
- 30：モニタ・ディスプレイ
- 32：音声出力バッファ
- 34：スピーカ

- 40：ワークステーション
- 42：JPEG圧縮伸長機能付きビデオ・ボード
- 110：ビデオ送信装置
- 112：大容量記憶装置
- 114：動画データ送信装置
- 116：音声データ送信装置
- 118：ネットワーク
- 120：ビデオ再生端末装置
- 122：動画データ受信装置
- 124：フレーム・バッファ
- 126：モニタ・ディスプレイ
- 128：音声データ受信装置
- 130：音声出力バッファ
- 132：スピーカ

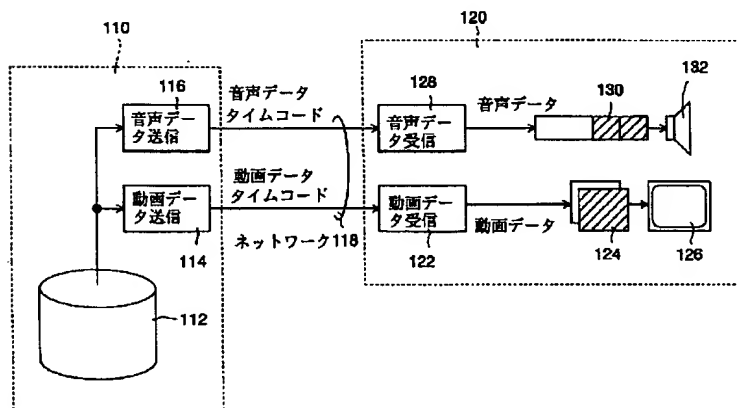
【図1】



【図4】



【図2】



【図3】

